

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
22 May 2003 (22.05.2003)

PCT

(10) International Publication Number
WO 03/042656 A2

- (51) International Patent Classification⁷: **G01N**
- (21) International Application Number: **PCT/US02/34879**
- (22) International Filing Date: 30 October 2002 (30.10.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/350,650 13 November 2001 (13.11.2001) US
- (71) Applicant (for all designated States except US): **GENENCOR INTERNATIONAL, INC.** [US/US]; 925 Page Mill Road, Palo Alto, CA 94304 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **HARDING, Fiona, A** [US/US]; 772 Lewis Street, Santa Clara, CA 95050 (US). **POWER, Scott, D.** [UG/UG]; 524 Poplar Avenue, San Bruno, CA 94066 (US).
- (74) Agent: **MACKNIGHT, Kamrin, T.**; GENENCOR INTERNATIONAL, INC., 925 Page Mill Road, Palo Alto, CA 94304 (US).

- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 03/042656 A2

(54) Title: IDENTIFYING AND REDUCING THE ALLERGENICITY OF FOOD PROTEINS

MAKISVAAAA LLVLMALGHA TAFRATVTTT VVEENQEEC REQMQRQOML
SHCRMYMRQQ MEESPYQTMP RRGMEPHMSE CCEQLEGMD E SCRCEGLRMM
MMRMQQEEMQ PRGEQMRMM RLAENIPSRC NLSPMRCPMG GSIAGF

(57) Abstract: The present invention provides methods and composition suitable for the identification and modification of allergenic proteins. More particularly, the present invention provides methods and compositions for the identification and modification of epitopes of allergenic food proteins, as well as epitopes of potentially allergenic food proteins.

IDENTIFYING AND REDUCING THE ALLERGENICITY OF FOOD PROTEINS

5

The present application claims priority to U.S. Provisional Patent Application Serial No. 60/350,650, filed on November 13, 2002.

10 Field of the Invention

The present invention provides methods and composition suitable for the identification and modification of allergenic proteins. More particularly, the present invention provides methods and compositions for the identification and modification of epitopes of allergenic food proteins, as well as epitopes of potentially allergenic food proteins.

15

Background of the Invention

Much work has been done to alleviate the problems of hypersensitivity in some individuals to proteins used in industrial, pharmaceutical and commercial applications.

20 However, efforts are also required to identify and potentially reduce the allergenicity of food proteins.

Food allergy is reported to be the most common cause of anaphylaxis outside of the hospital (Kemp *et al.*, Arch. Intern. Med., 1749 [1995]; and Yocum and Khan, Mayo Clinic Proc., 69:16-23 [1994]). Estimates indicate that from 5.4 million to 7 million Americans have
25 food allergies (Sicherer *et al.*, J. Allergy Clin. Immunol., 103:559-562 [2000]). Indeed, food allergies affect nearly everyone at some point. There are eight types of foods that are accountable for approximately 90% of all food-allergic reactions. The foods that most commonly cause anaphylaxis include peanuts, tree nuts, shellfish, fish, milk, soy, wheat and eggs. However, many other compounds have also been associated with food allergies. In
30 addition, with the on-going development of various genetically-modified crops, there is concern that the modifications to the plants will result in allergic reactions in those who eat them.

Currently, there are no good methods of preventing and treating food allergies. Indeed, prevention of anaphylaxis and less serious allergic reactions typically involves
35 avoidance of the foods that trigger the reaction. This is becoming increasingly difficult, as hidden ingredients may be included in food items, and product labels sometimes use

alternative names for food ingredients that contain allergens. In addition, food allergies in children are often difficult to manage due to adults or children inadvertently providing the problem food to a food-sensitive child.

Treatment of food allergies is likewise problematic and people with food allergies
5 need to always be prepared to treat their allergic reactions to food. Indeed, one study of children and adolescents with food allergies graphically demonstrates how important it is for those with food allergies to carry epinephrine with them at all times (Sampson *et al.*, New Eng. J. Med., 327:380-384 [1992]). According to this study, 10 out of 13 fatal and near-fatal anaphylactic reactions to food occurred in public places and none of those who died had
10 epinephrine with them. All of the cases who experienced such a reaction and received epinephrine before or within five minutes of developing severe symptoms survived. Thus, those who are highly sensitive to foods must carry epinephrine with them and know how to administer it to themselves. However, epinephrine use does not come without danger. Side effects of epinephrine include palpitations, tachycardia, sweating, nausea, vomiting, and
15 respiratory difficulty. Thus, even after administration of epinephrine, emergency medical treatment is often needed, in order to deal with any respiratory or cardiac emergency. Following the administration of epinephrine, sequelae such as cardiac arrhythmias sometimes occur. In addition to epinephrine, some patients need to carry anti-histamines with them at all times as well. Those who are severely allergic to certain food items must be
20 diligent in ensuring that they have their needed doses and means to administer the drug(s) to themselves. This is cumbersome and often impractical. Thus, what is needed are means to identify and modify food allergens such that their allergenicity is reduced. Indeed, the development of hypoallergenic food items would provide a great benefit to those who are sensitive to food allergens.

25

Summary of the Invention

The present invention provides methods and composition suitable for the identification and modification of allergenic proteins. More particularly, the present invention provides methods and compositions for the identification and modification of epitopes of
30 allergenic food proteins, as well as epitopes of potentially allergenic food proteins.

In some preferred embodiments, the present invention provides for identifying the allergenicity of a food protein comprising the steps of: obtaining a food protein; preparing a plurality of amino acid fragments of the food protein, such that each fragment overlaps in sequence with its contiguous fragments; contacting the amino acid fragments of the food
35 protein with a solution comprising naïve human CD4+ or CD8+ T-cells and dendritic cells, wherein the dendritic cells have been differentiated; and identifying an epitope region with

the amino acid fragments of the food protein, wherein the identifying step comprises measuring the ability of the epitope region to stimulate proliferation of the naïve human CD4+ or CD8+ T-cells. In some particularly preferred embodiments, the dendritic cells and the CD4+ or CD8+ cells are obtained from a single blood source. In alternative preferred
5 embodiments, the food protein is selected from the group consisting of nuts, seafood, dairy products, soy, wheat, eggs, and genetically-modified food products. In yet other embodiments, the epitope region comprises a T-cell epitope.

The present invention further provides methods for reducing the allergenicity of a food protein comprising the steps of: a) obtaining a food protein; b) preparing a plurality of
10 amino acid fragments of the food protein, such that each fragment overlaps in sequence with its contiguous fragments; c) contacting the amino acid fragments of the food protein with a first solution comprising naïve human CD4+ or CD8+ T-cells and dendritic cells, wherein the dendritic cells have been differentiated; and d) identifying an epitope region with
15 the amino acid fragments of the food protein, wherein the identifying comprises measuring the ability of the epitope region to stimulate proliferation of the naïve human CD4+ or CD8+ T-cells; and e) modifying the epitope region of the allergenic food protein identified in step d) to produce a food protein having reduced allergenicity. In some particularly preferred embodiments, the methods further comprise the step of determining the allergenicity of the food protein having reduced allergenicity.

20 In other embodiments of the methods, the additional steps of contacting a plurality of amino acid fragments of the food protein having reduced allergenicity with a second solution comprising naïve human CD4+ or CD8+ T-cells and dendritic cells, wherein the dendritic cells have been differentiated; measuring the ability of the amino acid fragments of the mutant protease to stimulate proliferation of the naïve human CD4+ or CD8+ T-cells; and
25 comparing the ability of the amino acid fragments of the food protein having reduced allergenicity to stimulate the proliferation of naïve human CD4+ or CD8+ T-cells with the ability of the amino acid fragments of the food allergen to stimulate the proliferation of naïve human CD4+ or CD8+ T-cells are provided. In still other embodiments, the dendritic cells and the CD4+ or CD8+ T-cells in the first and second solutions are obtained from a single
30 blood source. In additional embodiments, the epitope is a T-cell epitope. In further embodiments, the methods also the step of producing an expression vector comprising a nucleic acid sequence encoding the food allergen having reduced allergenicity. In still other embodiments, the methods further comprise the step of transforming at least one host cell with the expression vector. In yet additional embodiments, the methods further comprise
35 the steps of cultivating at least one host cell in a culture medium under conditions that promote the expression of the food protein having reduced allergenicity and recovering the

food allergen having reduced allergenicity from the cell or the culture medium.

Description of the Figures

5 Figure 1 illustrates the amino acid sequence of the mature protein of the Brazil nut seed storage protein known as allergen Ber e1 having SWISS-PROT accession number P04403. The mature protein Ber e1 protein (SEQ ID NO: 1) is made up of two small polypeptide chains: a heavy chain (73 amino acids) and a light chain (28 amino acids). The underlined sequences represent the epitope regions of the light and heavy chain proteins.

10 Figure 2 illustrates the protoxin amino acid sequence (SEQ ID NO: 2) of Cry 1Ab from *B. thuringiensis* serovar *aizawai* used to determine T-cell epitopes therein.

 Figure 3 illustrates the protoxin amino acid sequence (SEQ ID NO: 3) of Cry 3Aa from *B. thuringiensis* subsp. *San-diego* used to determine T-cell epitopes therein.

15 Figure 4 illustrates the results (percent responders) for each of the 15-mer peptide strings of Ber e1 protein. The light chain is represented by peptides 1-6 and the heavy chain is represented by peptides 7-27.

 Figure 5 illustrates the results (percent responders) for each of the 15-mer peptide strings of *Bacillus thuringiensis* cry protein Cry3Aa.

20 Figure 6 illustrates the results (percent responders) for each of the 15-mer peptide strings of the *Bacillus thuringiensis* cry protein Cry1Ab.

 Figure 7 is an alignment of the amino acid sequences for Cry1Ab and Cry3Aa

Definitions

25 Prior to providing a description of the invention, Applicants provide the following definitions.

30 As used herein, "allergenic food protein" refers to any food protein that is associated with causing an allergic reaction in humans and other animals. A "putative allergenic food protein" is a food protein that may be allergenic. A "food protein with reduced allergenicity" is a food protein that has been modified so as to be less allergenic (*i.e.*, hypoallergenic) than the original, unmodified protein. It is intended that these terms encompass naturally-occurring food proteins, as well as those produced synthetically and/or using recombinant technology.

"Antigen presenting cell" and "APC" as used herein refer to cells of the immune system that present antigen on their surface in a manner that is recognizable by T-cells. Examples of antigen presenting cells are dendritic cells, interdigitating cells, activated B-cells and macrophages.

5 "T-cell proliferation" as used herein means the number of T-cells produced during the incubation of T-cells with the antigen presenting cells, with or without antigen. In some preferred embodiments, T-cell proliferation is assessed by measuring the amount of tritiated thymidine incorporated into the DNA of dividing cells.

"Baseline T-cell proliferation" as used herein means T-cell proliferation that is
10 normally seen in an individual in response to exposure to antigen presenting cells in the absence of peptide or protein antigen. For the purposes herein, the baseline T-cell proliferation level was determined on a per sample basis for each individual as the proliferation of T-cells in response to antigen presenting cells in the absence of antigen.

The term "lymphoid" when used in reference to a cell line or a cell, means that the
15 cell line or cell is derived from the lymphoid lineage and includes cells of both the B and the T lymphocyte lineages.

The terms "T lymphocyte" and "T-cell" as used herein encompass any cell within the T lymphocyte lineage from T-cell precursors (including Thy1 positive cells which have not rearranged the T cell receptor genes) to mature T cells (*i.e.*, single positive for either CD4 or
20 CD8, surface TCR positive cells).

The terms "B lymphocyte" and "B-cell" encompasses any cell within the B-cell lineage from B-cell precursors, such as pre-B-cells (B220⁺ cells which have begun to rearrange Ig heavy chain genes), to mature B-cells and plasma cells.

As used herein, "CD4⁺ T-cell" and "CD4 T-cell" refer to helper T-cells, while "CD8⁺ T-cell" and "CD8 T-cell" refer to cytotoxic T-cells.
25

As used herein "T-cell epitope" means a feature of a peptide or protein that is recognized by a T-cell receptor in the initiation of an immunologic response to the peptide comprising that antigen. Recognition of a T-cell epitope by a T-cell is generally believed to be via a mechanism wherein T-cells recognize peptide fragments of antigens which are
30 bound to class I or class II Major Histocompatibility Complex (MHC) molecules expressed on antigen-presenting cells (See *e.g.*, Moeller (ed.), Immunol. Rev., 98:187 [1987]). In some embodiments of the present invention, the epitopes or epitopic fragments identified as described herein find use in the detection of antigen presenting cells having MHC molecules

capable of binding and displaying the epitopes or fragments. In some embodiments, the epitopes/epitopic fragments further comprise a detectable label (*i.e.*, a marker) that facilitates the identification of cells that bind and/or display the epitope/epitopic fragment of interest.

5 As used herein, a "major epitope" is defined as a peptide sequence that induces responses in a percentage of the donor pool equal to approximately three times the background rate.

The term "recombinant DNA molecule" as used herein refers to a DNA molecule that is comprised of segments of DNA joined together by means of molecular biological
10 techniques.

DNA molecules are said to have "5' ends" and "3' ends" because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage. Therefore, an end of an oligonucleotides is referred to as the "5'
15 end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide, also may be said to have 5' and 3' ends. In either a linear or circular DNA molecule, discrete elements are referred to as being "upstream" or 5' of the
20 "downstream" or 3' elements. This terminology reflects the fact that transcription proceeds in a 5' to 3' fashion along the DNA strand. The promoter and enhancer elements which direct transcription of a linked gene are generally located 5' or upstream of the coding region (enhancer elements can exert their effect even when located 3' of the promoter element and the coding region). Transcription termination and polyadenylation signals are located 3' or
25 downstream of the coding region.

The term "an oligonucleotide having a nucleotide sequence encoding a gene" means a DNA sequence comprising the coding region of a gene or, in other words, the DNA sequence that encodes a gene product. The coding region may be present in either a cDNA or genomic DNA form. Suitable control elements such as enhancers/promoters, splice
30 junctions, polyadenylation signals, etc. may be placed in close proximity to the coding region of the gene if needed to permit proper initiation of transcription and/or correct processing of the primary RNA transcript. Alternatively, the coding region utilized in the expression vectors of the present invention may contain endogenous enhancers/promoters, splice junctions, intervening sequences, polyadenylation signals, etc. or a combination of both

endogenous and exogenous control elements.

The term "recombinant oligonucleotide" refers to an oligonucleotide created using molecular biological manipulations, including but not limited to, the ligation of two or more oligonucleotide sequences generated by restriction enzyme digestion of a polynucleotide
5 sequence, the synthesis of oligonucleotides (e.g., the synthesis of primers or oligonucleotides) and the like.

The term "expression vector" as used herein refers to a recombinant DNA molecule containing a desired coding sequence and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host organism.
10 Nucleic acid sequences necessary for expression in prokaryotes include a promoter, optionally an operator sequence, a ribosome binding site and possibly other sequences. Eukaryotic cells are known to utilize promoters, enhancers, and termination and polyadenylation signals. Once transformed into a suitable host, the vector may replicate and function independently of the host genome, or may, in some instances, integrate into
15 the genome itself. As used herein, "plasmid" and "vector" are sometimes used interchangeably as the plasmid is the most commonly used form of vector at present. However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which are, or become, known in the art, including but not limited to plasmids, phage particles, or simply potential genomic inserts.

20 The "host cells" used in the present invention generally are prokaryotic or eukaryotic hosts which contain the expression vector and/or gene of interest. Host cells are transformed or transfected with vectors constructed using recombinant DNA techniques. Such transformed host cells are capable of either replicating vectors encoding the protein variants or expressing the desired protein variant. In the case of vectors which encode the
25 pre- or prepro-form of the protein variant, such variants, when expressed, are typically secreted from the host cell into the host cell medium.

Efficient expression of recombinant DNA sequences in eukaryotic cells requires signals directing the efficient termination and polyadenylation of the resulting transcript. Transcription termination signals are generally found downstream of the polyadenylation
30 signal and are a few hundred nucleotides in length. The term "poly A site" or "poly A sequence" as used herein denotes a DNA sequence which directs both the termination and polyadenylation of the nascent RNA transcript. Efficient polyadenylation of the recombinant transcript is desirable as transcripts lacking a poly A tail are unstable and are rapidly degraded. The poly A signal utilized in an expression vector may be "heterologous" or

"endogenous." An endogenous poly A signal is one that is found naturally at the 3' end of the coding region of a given gene in the genome. A heterologous poly A signal is one which is isolated from one gene and placed 3' of another gene. A commonly used heterologous poly A signal is the SV40 poly A signal.

5 The term "stable transfection" or "stably transfected" refers to the introduction and integration of foreign DNA into the genome of the transfected cell. The term "stable transfectant" refers to a cell which has stably integrated foreign or exogenous DNA into the genomic DNA of the transfected cell.

 The terms "selectable marker" or "selectable gene product" as used herein refer to
10 the use of a gene which encodes an enzymatic activity that confers resistance to an antibiotic or drug upon the cell in which the selectable marker is expressed. Selectable markers may be "dominant"; a dominant selectable marker encodes an enzymatic activity which can be detected in any mammalian cell line. Examples of dominant selectable
15 markers include the bacterial aminoglycoside 3' phosphotransferase gene (also referred to as the *neo* gene) which confers resistance to the drug G418 in mammalian cells, the bacterial hygromycin G phosphotransferase (*hyg*) gene which confers resistance to the antibiotic hygromycin and the bacterial xanthine-guanine phosphoribosyl transferase gene (also referred to as the *gpt* gene) which confers the ability to grow in the presence of mycophenolic acid. Other selectable markers are not dominant in that their use must be in
20 conjunction with a cell line that lacks the relevant enzyme activity. Examples of non-dominant selectable markers include the thymidine kinase (*tk*) gene which is used in conjunction with TK⁻ cell lines, the carbamoyl-phosphate synthetase-aspartate transcarbamoylase-dihydroorotase (CAD) gene which is used in conjunction with CAD-deficient cells and the mammalian hypoxanthine-guanine phosphoribosyl transferase (*hprt*)
25 gene which is used in conjunction with HPRT⁻ cell lines. It is noted that some selectable markers can be amplified and therefore can be used as amplifiable markers (e.g., the CAD gene).

 The terms "nucleic acid molecule encoding," "DNA sequence encoding," and "DNA encoding" refer to the order or sequence of deoxyribonucleotides along a strand of
30 deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along the polypeptide (protein) chain. The DNA sequence thus codes for the amino acid sequence.

 "Homolog" as used herein means a protein (e.g., an enzyme) which has similar catalytic action, structure and/or use as the protein of interest. It is desirable to find a

homolog that has a tertiary and/or primary structure similar to the protein of interest as replacement of the epitope in the protein of interest with an analogous segment from the homolog will reduce the disruptiveness of the change. Thus, in most cases, closely homologous proteins provide the most desirable source of epitope substitutions.

- 5 Alternatively, if possible, it is advantageous to look to human analogs for a given protein. For example, in some embodiments, substituting a specific epitope in one food allergen with a sequence from another food protein that is not allergenic results in the production of a food protein having reduced allergenicity.

As used herein, the term "substantially identical" in the context of two nucleic acids
10 or polypeptides thus typically means that a polynucleotide or polypeptide comprises a sequence that has at least 70% sequence identity, preferably at least 80%, more preferably at least 90%, still more preferably 95%, most preferably 97%, and sometimes as much as 97%, compared to the reference sequence using the known programs such as BLAST, ALIGN, and CLUSTAL using standard parameters (See *e.g.*, Altschul *et al.*, J. Mol. Biol.
15 215:403-410 [1990]; See also, Henikoff *et al.*, Proc. Natl. Acad. Sci. USA 89:10915 [1989]; and Karin *et al.*, Proc. Natl. Acad. Sci. USA 90:5873 [1993]). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. One indication that two polypeptides are substantially identical is that the first polypeptide is immunologically cross-reactive with the second polypeptide. Typically, polypeptides that
20 differ by conservative amino acid substitutions are immunologically cross-reactive. Thus, for example, a polypeptide is substantially identical to a second polypeptide, where the two peptides differ only by a conservative substitution. Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions (*e.g.*, within a range of medium to high stringency).

25 As used herein, the term "analogous" sequence refers to a sequence that shares particular characteristics with the original protein of interest. Analogous sequences may be determined by ensuring that the replacement amino acids show a similar function, the tertiary structure and/or conserved residues to the amino acids in the protein of interest at or near the epitope. Thus, where the epitope region contains, for example, an alpha-helix or a
30 beta-sheet structure, the replacement amino acids preferably maintain that specific structure.

As used herein, "equivalent residues" refers to proteins that share particular amino acid residues. For example, equivalent residues may be identified by determining homology at the level of tertiary structure for a protein (*e.g.*, a food allergen such as Ber e1) whose
35 tertiary structure has been determined by x-ray crystallography. Equivalent residues are

defined as those for which the atomic coordinates of two or more of the main chain atoms of a particular amino acid residue of the protein having putative equivalent residues and the protein of interest (N on N, CA on CA, C on C and O on O) are within 0.13 nm and preferably 0.1 nm after alignment. Alignment is achieved after the best model has been
5 oriented and positioned to give the maximum overlap of atomic coordinates of non-hydrogen protein atoms of the proteins analyzed. The preferred model is the crystallographic model giving the lowest R factor for experimental diffraction data at the highest resolution available, determined using methods known to those skilled in the art of crystallography and protein characterization/analysis.

10 Some of the residues identified for substitution, insertion or deletion are conserved residues whereas others are not. In the case of residues which are not conserved, the replacement of one or more amino acids is limited to substitutions which produce a variant which has an amino acid sequence that does not correspond to one found in nature. In the case of conserved residues, such replacements should not result in a naturally-occurring
15 sequence.

In some embodiments, the following cassette mutagenesis method finds use in facilitating the construction of the protein variants of the present invention, although other methods may be used. First, the naturally-occurring gene encoding the protein is obtained and sequenced in whole or in part. Then the sequence is scanned for a point at which it is
20 desired to make a mutation (deletion, insertion or substitution) of one or more amino acids in the encoded protein. The sequences flanking this point are evaluated for the presence of restriction sites for replacing a short segment of the gene with an oligonucleotide pool which when expressed will encode various mutants. Such restriction sites are preferably unique sites within the protein gene so as to facilitate the replacement of the gene segment.
25 However, any convenient restriction site which is not overly redundant in the protein gene may be used, provided the gene fragments generated by restriction digestion can be reassembled in proper sequence. If restriction sites are not present at locations within a convenient distance from the selected point (from 10 to 15 nucleotides), such sites are generated by substituting nucleotides in the gene in such a fashion that neither the reading
30 frame nor the amino acids encoded are changed in the final construction. Mutation of the gene in order to change its sequence to conform to the desired sequence is accomplished by M13 primer extension in accord with generally known methods. The task of locating suitable flanking regions and evaluating the needed changes to arrive at two convenient restriction site sequences is made routine by the redundancy of the genetic code, a
35 restriction enzyme map of the gene and the large number of different restriction enzymes.

Note that if a convenient flanking restriction site is available, the above method need be used only in connection with the flanking region which does not contain a site.

Once the naturally-occurring DNA or synthetic DNA is cloned, the restriction sites flanking the positions to be mutated are digested with the cognate restriction enzymes and a plurality of end termini-complementary oligonucleotide cassettes are ligated into the gene. The mutagenesis is simplified by this method because all of the oligonucleotides can be synthesized so as to have the same restriction sites, and no synthetic linkers are necessary to create the restriction sites.

10 Description of the Invention

The present invention provides methods and composition suitable for the identification and modification of allergenic proteins. More particularly, the present invention provides methods and compositions for the identification and modification of epitopes of allergenic food proteins, as well as epitopes of potentially allergenic food proteins.

15 The present invention further provides the CD4+ T-cell epitopes of the 2S globulin storage protein of the Brazil nut (*Bertholletia excelsa*). The allergen associated with this protein is designated "Ber e1" (See e.g., Nordlee *et al.*, New Eng. J. Med., 334:688-692 [1996]). The polynucleotide sequence and protein sequence of this protein are provided in Altenbach *et al.*, Plant Mol. Biol., 8:239-250 (1987); Gander *et al.*, Plant Mol. Biol. 16:437-
20 448 (1991); and Ampe *et al.*, Eur. J. Biochem. (FEBS) 159:597-604 (1986).

In addition, the present invention provides the CD4+ T-cell epitopes of two known *Bacillus thuringiensis* Cry proteins. *Bacillus thuringiensis* (Bt) is a bacterial species that produces a crystal protein that is a protoxin. This protoxin is also referred to as a delta-endotoxin. The protoxin is converted into a toxin upon ingestion by insect larvae. The toxin
25 is particularly effective against lepidopteran, coleopteran and dipteran insects that create major problems in agriculture and forestry. Many Bt proteins have been isolated and sequenced. These include but are not limited to Cry1Aa, Cry1Ab, Cry1Ac, CryIC, CryIE, CryICb, CryIF, Cry3Aa, Cry5 and others. Various classifications exist for the Cry proteins. For example, Hofte and Whiteley, Microbiol. Rev., 53:242-255 (1989) present one early
30 example of a classification system, but many new Cry proteins have been discovered and isolated since this classification was proposed.

A number of genes have been constructed which include the coding sequence for Bt Cry proteins. In particular, corn and potato cells have been transformed with constructs

including Bt cry genes or genes encoding a protein having substantially the insect toxicity properties of the Bt toxin, in order to increase the productivity of these crops. In contrast to Brazil nut which may cause a severe allergenic response in certain humans, the Cry proteins which have been incorporated into animal and human crop plants is not known to
5 cause any significant allergenic response. However, there is concern that genetically-modified foods will result in allergic responses in humans and other animals due to the introduction of proteins not previously in the food supply. Thus, means to identify and/or modify food proteins that may be allergenic find use in assessing the safety of genetically-modified foods.

10 Indeed, the present invention provides means to determine the T-cell epitopes in such proteins commonly used in transgenic food crops. Cry1Ab is one of the most common Cry proteins and shares at least 80% amino acid sequence homology with CryIAa and CryIAc (See *e.g.*, Hofte *et al.*, Eur. J. Biochem. 161:273 [1986]). Thus, the present invention also provides means to identify and/or modify potential T-cell epitopes in Cry3Aa (See *e.g.*,
15 Hofte *et al.*, Nucl. Acids Res., 15: 7183 [1987]; and Sekar *et al.*, Proc. Natl. Acad. Sci. USA 847036-7040 [1987]).

In some particularly preferred embodiments, epitopes are identified using an assay system that identifies epitopes and non-epitopes. In these embodiments, differentiated dendritic cells are combined with human CD4+ and/or CD8+ T-cells and with a peptide of
20 interest (*e.g.*, a food allergen or putative food allergen). More specifically, a food protein having reduced allergenicity can be provided. In this assay system, a T-cell epitope is recognized by means comprising the steps of: (a) obtaining from a single blood source a solution of dendritic cells and a solution of naïve CD4+ and/or CD8+ T-cells; (b) promoting differentiation in said solution of dendritic cells; (c) combining said solution of differentiated
25 dendritic cells and said CD4+ and/or CD8+ T-cells with a peptide of interest; (d) measuring the proliferation of the T-cells in said step (c).

In some embodiments of the invention, a series of peptide oligomers that correspond to all or parts of the test protein (*e.g.*, food allergen) are prepared. For example, a peptide library is produced covering the relevant portion or all of the test protein. In one
30 embodiment, the manner of producing the peptides is to introduce overlap into the peptide library, for example, producing a first peptide corresponds to amino acid sequence 1-15 of the test protein, a second peptide corresponds to amino acid sequence 4-18 of the test protein, a third peptide corresponds to amino acid sequence 7-21 of the test protein, a fourth peptide corresponds to amino acid sequence 10-24 of the test protein etc., until
35 representative peptides corresponding to the entire molecule are created. By analyzing each of the peptides individually in the assay provided herein, it is possible to precisely

identify the location of epitopes recognized by T-cells. In the example above, the greater reaction of one specific peptide than its neighbors' facilitates identification of the epitope anchor region to within three amino acids. After determining the location of these epitopes, it is possible to alter the amino acids within each epitope until the peptide produces a different T-cell response from that of the original protein. Moreover, food proteins may be identified herein which have desired low T-cell epitope potency which may be used in their naturally occurring forms.

Various *in vitro* and *in vivo* assays known in the art find use in the ascertainment of the reduced immunogenic response of variant proteins. Suitable assay systems include, but are not limited to *in vivo* assays such as HLA-DR3/DQ2 mouse T cell responses and *in vitro* assays such as human peripheral blood mononuclear cells (PBMC) assays (See, Herman *et al.*, J. Immunol., 163:6275-6282; Sonderstrup *et al.*, Immunol. Rev., 172: 335-343 [1999]; Taneja and David, Immunol. Rev., 169:67-79 [1999]; and Grusby *et al.*, Proc. Natl. Acad. Sci. USA 90:3913-3917 [1993]).

In one embodiment of the present invention, one or more CD4+ T-cell epitopes were determined for each of the Ber e1 protein, Cry 1Ab protein and Cry3Aa protein.

In another embodiment the identified T-cell epitope is modified to produce variant(s) with reduced allergenic capacity. In a further embodiment, the present invention provides means to reduce the allergenic response of a Ber e1 protein comprising obtaining a precursor Ber e1 protein and obtaining a variant of the precursor Ber e1 protein, wherein the variant has at least one T-cell epitope of the precursor protein and wherein the variant exhibits an altered allergenic response which differs from the allergenic response of the precursor Ber e1 protein.

In yet a further embodiment of the present invention, the epitopes identified herein can be used to elicit an immune response (e.g., where it is desired to raise antibodies against a protein of interest). Such antibodies can be used, for example, to screen for other food proteins that include one or more of the epitope regions, or regions highly homologous thereto. Thus, the methods of the present invention find use in the development of additional systems for use in assessing food allergies.

Additionally, the epitopic fragments herein can be used in the detection of antigen presenting cells having MHC molecules capable of binding and displaying such fragments. For example, the epitopic fragments can include a detectable label (e.g., radiolabel). The labeled fragments can then be incubated with cells of interest, and then cells that bind (or display) the labeled fragments can be detected.

In additional embodiments, the present invention provides means to determine the HLA associations with reactivity to specific food allergen epitopes. For example in experiments conducted during the development of the present invention, the HLA-DR and DQ types of donors who responded to peptide #4 (light chain amino acids 10-24 of Ber e1) and peptide #30 (heavy chain amino acids 40-55 of Ber e1) were analyzed for the presence of any statistically significant enrichments. HLA-DQ2 was found to be significantly associated with a non-response to peptide #4 ($p = 0.05$, relative risk = 0.29). There was a positive association between the presence of HLA-DR15(2) and a response to peptide #20 ($p = 0.04$, relative risk = 3.0). The strongest association found was a positive association between a response to peptide #12 and the presence of HLA-DR13 ($p = 0.01$, relative risk = 4.44). Thus, it is contemplated that the present invention will find use in identifying individuals who are potentially allergic to any food allergen of interest.

EXPERIMENTAL

The following examples are provided in order to demonstrate and further illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

In the experimental disclosure which follows, the following abbreviations apply: °C (degrees Centigrade); rpm (revolutions per minute); xg (times gravity); IgG (immunoglobulin G); IM (intramuscular); IP (intraperitoneal); IV (intravenous or intravascular); SC (subcutaneous); H₂O (water); HCl (hydrochloric acid); aa (amino acid); bp (base pair); kb (kilobase pair); kD (kilodaltons); gm (grams); µg (micrograms); mg (milligrams); ng (nanograms); µl (microliters); ml (milliliters); mm (millimeters); nm (nanometers); µm (micrometer); M (molar); mM (millimolar); µM (micromolar); U (units); V (volts); MW (molecular weight); sec (seconds); min(s) (minute/minutes); hr(s) (hour/hours); MgCl₂ (magnesium chloride); NaCl (sodium chloride); OD₂₈₀ (optical density at 280 nm); OD₆₀₀ (optical density at 600 nm); PAGE (polyacrylamide gel electrophoresis); PBS (phosphate buffered saline [150 mM NaCl, 10 mM sodium phosphate buffer, pH 7.2]); PCR (polymerase chain reaction); RT-PCR (reverse transcription PCR); SDS (sodium dodecyl sulfate); Tris (tris(hydroxymethyl)aminomethane); w/v (weight to volume); v/v (volume to volume); Amersham (Amersham Life Science, Inc. Arlington Heights, IL); ICN (ICN Pharmaceuticals, Inc., Costa Mesa, CA); Amicon (Amicon, Inc., Beverly, MA); Clontech (CLONTECH Laboratories, Palo Alto, CA); Difco (Difco Laboratories, Detroit, MI); GIBCO BRL or Gibco BRL (Life Technologies, Inc., Gaithersburg, MD); Invitrogen (Invitrogen Corp., San Diego, CA); Endogen (Endogen, Woburn, MA); CytoVax (CytoVax, Edmonton, Canada); Wyeth-

Ayerst (Wyeth-Ayerst, Philadelphia, PA); NEN (NEN Life Science Products, Boston, MA); Wallace Oy (Wallace Oy, Turku, Finland); Pharma AS (Pharma AS, Oslo, Norway); Dynal (Dynal, Oslo, Norway); Bio-Synthesis (Bio-Synthesis, Lewisville, TX); ATCC (American Type Culture Collection, Rockville, MD); Gibco/BRL (Gibco/BRL, Grand Island, NY); Sigma (Sigma Chemical Co., St. Louis, MO); Pharmacia (Pharmacia Biotech, Pisacataway, NJ); and Stratagene (Stratagene, La Jolla, CA).

EXAMPLE 1

Assay for the Identification of Peptide T-Cell Epitopes in Brazil Nut Seed Storage Protein and in *Bacillus thuringiensis* Cry Protein Using Human T-Cells

Fresh human peripheral blood cells were collected from 48, 47 or 92 humans of unknown exposure status to food proteins and were tested for the determination of antigenic epitopes in Cry1Ab, Cry3Aa, and Ber e1. Peripheral mononuclear blood cells (stored at room temperature, no older than 24 hours) were prepared for use as described herein. Approximately 30 mls of a solution of buffy coat preparation from one unit of whole blood was brought to 50 ml with Dulbecco's phosphate buffered solution (DPBS) and split into two tubes. The samples were underlaid with 12.5 ml of room temperature lymphoprep density separation media (Nycomed density 1.077 g/ml). The tubes were centrifuged for thirty minutes at 600 xg. The interface of the two phases was collected, pooled and washed in DPBS. The cell density of the resultant solution was measured by hemocytometer, as known in the art. Cell viability was measured by trypan blue exclusion, as known in the art.

From the resulting solution, a differentiated dendritic cell culture was prepared from the peripheral blood mononuclear cell sample having a density of 10^8 cells per 75 ml culture flask in a solution as follows:

- (1) 50 ml of serum free AIM V medium (Gibco) was supplemented with a 1:100 dilution beta-mercaptoethanol (Gibco). The flasks were laid flat for two hours at 37°C in 5% CO₂ to allow adherence of monocytes to the flask wall;
- (2) Differentiation of the monocyte cells to dendritic cells was performed as follows: nonadherent cells were removed and the resultant adherent cells (monocytes) combined with 30 ml of AIM V, 800 units/ml of GM-CSF (Endogen) and 500 units/ml of IL-4 (Endogen); the resulting mixture was cultured for 5 days under conditions at 37°C in 5% CO₂. After five days, the cytokine TNFalpha (Endogen)

was added to 0.2 units/ml, and the cytokine IL-1a (Endogen) was added to a final concentration of 50 units/ml and the mixture incubated at 37°C in 5% CO₂ for two more days;

5 (3) On the seventh day, mitomycin C was added to a concentration of 50 microgram/ml was added to stop growth of the now differentiated dendritic cell culture. The solution was incubated for 60 minutes at 37°C in 5% CO₂. Dendritic cells were collected by gently scraping the adherent cells off the bottom of the flask with a cell scraper. Adherent and non-adherent cells were then centrifuged at 600xg for 5 minutes, washed in DPBS and counted; and

10 (4) The prepared dendritic cells were placed into a 96 well round bottom array plate at 2×10^4 cells/well in 100 microliter total volume of AIM V medium.

CD4+ T cells were prepared from frozen aliquots of the peripheral blood cell samples used to prepare the dendritic cells. These T-cells were prepared using the human CD4+ 15 Collect Kit (Biotex) as per the manufacturer's instructions, but with the following modifications: the aliquots were thawed and washed such that approximately 10^8 cells were applied to each Collect column. First, the cells were suspended in 4 ml DPBS and 1 ml of the Cell reagent from the Collect Kit, with the solution being maintained at room temperature for 20 minutes. Following this 20 minute incubation, the cell solution was centrifuged for five 20 minutes at 600xg at room temperature and the pellet was resuspended in 2 ml of DPBS and applied to the Collect columns. The effluent from the columns was collected in 2% human serum in DPBS. The resultant CD4+ cell solution was centrifuged, resuspended in AIMV medium and the cell density determined. The CD4+ T-cell suspension was adjusted to provide 2×10^6 cells/ml in AIM V medium to facilitate efficient manipulation of the 96 well plate 25 used in the assay system.

Synthetic 15-mer peptides corresponding to the sequence of the test protein are resuspended in DMSO to create a stock solution at a concentration of 1 mM. Peptide antigen was prepared for use in the assay by diluting the 1mM stock into AIM V media at a ratio of 1:10. Then, 10 µl of the diluted peptide solution was added to each well of the 96 30 well plate containing the differentiated dendritic cells. Finally, 100 microliters of the diluted CD4+ T-cell solution as prepared above was added to each well. Useful controls include diluted DMSO blanks, and tetanus toxoid positive controls.

The final concentrations in each well, at 210 microliter total volume are as follows:

2×10^4 CD4+ cells

35 2×10^5 dendritic cells (R:S of 10:1)

5 µM peptide

Example 2**Identification of T-Cell Epitopes in Ber e1 from *Bertholletia excelsa*
and from Cry 1Ab and Cry3A from *Bacillus thuringiensis***

5 Peptides for use in the assay described in Example 1 were prepared based on the Ber e1 protein of *Bertholletia excelsa*, as well as Cry 1Ab, and Cry3Aa from *Bacillus thuringiensis* amino acid sequences (See, Figures 1, 2, and 3). Peptide antigens were designed essentially the same for each protein. From the full length amino acid sequence of the protein of interest starting with the first amino acid residue as provided in Figures 1 - 3,
10 15mers were synthetically prepared, with each 15mer overlapping with the previous and the subsequent 15mer except for three residues. All tests were performed at least in duplicate. All tests reported displayed robust positive control responses to the antigen tetanus toxoid. Responses were averaged within each experiment, then normalized to the baseline response. A positive event was recorded if the response was at least 2.95 times the
15 baseline response.

 The results are illustrated in Figures 4, 5 and 6. Results for all donors were compiled. The percentage of the donor pool that responded with a baseline-normalized response of 2.95 or better was calculated for each peptide. The background response rate for each peptide set corresponding to each protein tested was calculated by averaging all the percent
20 responses to all the peptides within a set. The background response for the Ber e 1 peptide set was determined to be 4.27%. As indicated above, a "major epitope" is a peptide sequence that induces responses in a percentage of the donor pool equal to three times the background rate. For the Ber e 1 dataset, this is equal to a value of 12.81%. As shown in Figure 4, only the response to peptide #4 reached the three-fold cut-off (peptide #4 value =
25 18.48%). Weak epitopes are designated by responses that reach or exceed twice the background rate (*i.e.*, 8.54% for the Ber e 1 dataset). Two peptides (#12 and #20) were identified as minor epitopes. Peptide # 16 is prominent in the dataset, but does not reach any epitope designation. These peptides correspond to the following sequences:

30	Peptide 4 = RQQMLSHCRMVYMRQQ	(SEQ ID NO: 7)
	Peptide 12 = LEGMDESCRCEGLRM	(SEQ ID NO: 4)
	Peptide 16 = LRMMMMRMQQEEMQP	(SEQ ID NO: 5)
	Peptide 20 = MQPRGEQMRRMMRLA	(SEQ ID NO: 6)

 As indicated in Figure 6, only 1 major T-cell epitope was determined from Cry1Ab.
35 The background value was calculated for the dataset in the absence of the zero values. The major epitope (Peptide 145; VSMFRSGFSNSSVSI [SEQ ID NO: 8]) in this protein reached

three times this background rate. However, multiple weak epitopes for Cry1Ab were observed, including:

- Peptide 26 = FLVQIEQLINQRIEE (SEQ ID NO: 9)
Peptide 35 = LSNLYQIYAESFREW (SEQ ID NO: 10)
5 Peptide 124 = PFNIGINNQQLSVLD (SEQ ID NO: 11)
Peptide 137 = SLDEIPPQNNNVPPR (SEQ ID NO: 12)
Peptide 174 = SQRYRVRIRYASTTN (SEQ ID NO: 13)

As indicated in Figure 5, no major T-cell epitopes were determined for Cry3Aa.

10 However, two weak epitopes were observed:

- Peptide 40 = DQKIADYAKNKALAE (SEQ ID NO: 14)
Peptide 47 = NVEDYVSALSSWQKN (SEQ ID NO: 15)

15

Example 3

Determination of Variants with Altered Allergenicity Residues

Within an Epitope

20 Peptide variants based on peptide # 12, 16 and 20 are tested in the mapping assay described above, with 20 or more community donor blood samples. For each of the peptide variants, three amino acid offset 15-mers are constructed to cover the entire region of the proposed change. The parent peptides in the set may be analyzed by mass spectrometry, as desired. In addition, crystal structure analysis is sometimes performed, in order to guide the selection of amino acid variants, as this helps preserve the function and structure of the
25 molecule.

An alanine scan is performed for each peptide. In addition, variant peptides corresponding to the major and weak epitopes are generated using methods known in the art. The assay is performed as described in Example 3, utilizing the variant peptides on a set of donor samples. Proliferative responses are collated, and the results compared with
30 the results obtained with the original peptide sequences. In some experiments, additional testing is used to confirm the statistical relevance of the results. Based on the known structure of the food allergens, a number of alternative sequences in the peptides corresponding to the major and weak epitopes (as desired) are selected, such that they are predicted to have little effect on the protein structure, but are different enough to
35 potentially affect the recognition of the epitope by the immune system. These modified epitopes then find use in modifying food allergens such that the food allergens are

hypoallergenic.

From the above, it is clear that the present invention provides methods and compositions for the identification of T-cell epitopes in food allergen proteins and the production of peptides which when incorporated into the food allergen sequence, are no longer capable of initiating the CD4⁺ T-cell response. In particular, the present invention provides means and compositions suitable for reducing the allergenicity of food proteins.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology, immunology, and/or related fields are intended to be within the scope of the present invention.

CLAIMS

1. A method for identifying the allergenicity of a food protein comprising the
5 steps of:
- a) obtaining the amino acid sequence of a food protein;
 - b) preparing a plurality of amino acid fragments of said food protein,
such that each fragment overlaps in sequence with its contiguous fragments;
 - c) contacting said amino acid fragments of said food protein with a
10 solution comprising naïve human CD4+ or CD8+ T-cells and dendritic cells, wherein
said dendritic cells have been differentiated; and
 - d) identifying an epitope region with said amino acid fragments of said
food protein, wherein said identifying comprises measuring the ability of said epitope
region to stimulate proliferation of said naïve human CD4+ or CD8+ T-cells.
- 15
2. The method of Claim 1, wherein said dendritic cells and said CD4+ or CD8+
cells are obtained from a single blood source.
3. The method of Claim 1, wherein said food protein is selected from the group
20 consisting of nuts, seafood, dairy products, soy, wheat, eggs, and genetically-modified food
products.
4. The method of Claim 1, wherein said epitope region comprises a T-cell
epitope.
- 25
5. A method for reducing the allergenicity of a food protein comprising the steps
of:
- a) obtaining the amino acid sequence of a food protein;
 - b) preparing a plurality of amino acid fragments of said food protein,
30 such that each fragment overlaps in sequence with its contiguous fragments;
 - c) contacting said amino acid fragments of said food protein with a first
solution comprising naïve human CD4+ or CD8+ T-cells and dendritic cells, wherein
said dendritic cells have been differentiated; and
 - d) identifying an epitope region with said amino acid fragments of said
35 food protein, wherein said identifying comprises measuring the ability of said epitope
region to stimulate proliferation of said naïve human CD4+ or CD8+ T-cells;

e) modifying said epitope region of said allergenic food protein identified in step d) to produce a food protein having reduced allergenicity.

6. The method of Claim 5, further comprising the step of determining the
5 allergenicity of said food protein having reduced allergenicity.

7. The method of Claim 6, further comprising the steps of:

f) contacting a plurality of amino acid fragments of said food protein
having reduced allergenicity with a second solution comprising naïve human CD4+ or
10 CD8+ T-cells and dendritic cells, wherein said dendritic cells have been
differentiated;

g) measuring the ability of said amino acid fragments of said mutant
protease to stimulate proliferation of said naïve human CD4+ or CD8+ T-cells; and

h) comparing the ability of said amino acid fragments of said food protein
15 having reduced allergenicity to stimulate the proliferation of naïve human CD4+ or
CD8+ T-cells with the ability of said amino acid fragments of said food allergen to
stimulate the proliferation of naïve human CD4+ or CD8+ T-cells.

8. The method of Claim 6, wherein said dendritic cells and said CD4+ or CD8+
20 T-cells in said first and second solutions are obtained from a single blood source.

9. The method of Claim 6, wherein said epitope is a T-cell epitope.

10. The method of Claim 6, further comprising the step of producing an
25 expression vector comprising a nucleic acid sequence encoding said food allergen having
reduced allergenicity.

11. The method of Claim 10, further comprising the step of transforming at least
one host cell with said expression vector.

30

12. The method of Claim 11, further comprising the steps of cultivating said at
least one host cell in a culture medium under conditions that promote the expression of said
food protein having reduced allergenicity and recovering said food allergen having reduced
allergenicity from said cell or said culture medium.

35

1/3

FIGURE 1

MAKISVAAAA LLVLMALGHA TAFRATVTTT VVEENQEEC REQMQRQOML
SHCRMYMRQQ MEESPYQTMP RRGMEPHMSE CCEQLEGMDE SCRCEGLRMM
MMRMQQEEMQ PRGEQMRMM RLAENIPSRC NLSPMRCPMG GSIAGF

FIGURE 2

MDNNPNINEC IPYNCLSNPE VEVLGGERIE TGYTPIDISL SLTQFLLSEF
VPGAGFVLGL VDIIWGIFGP SQWDAFLVQI EQLINQRIEE FARNQAISRL
EGLSNLYQIY AESFREWEAD PTNPALREEM RIQFNDMNSA LTTAIPLFAV
QNYQVPLLSV YVQAANLHLS VLRDVSFVGQ RWGFDAATIN SRYNDLTRLI
GNYTDHAVRW YNTGLERVWG PDSRDWIRYN QFRRELTTLV LDIVSLFPNY
DSRTYPIRTV SQLTREIYTN PVLENFDGSF RGSAQGIEGS IRSPHLM DIL
NSITIYTDAH RGEYYWSGHQ IMASPVGFSG PEFTFPLYGT MGNAAPQORI
VAQLQGQVYR TLSSTLYRRP FNIGINNQQL SVLDGTEFAY GTSSNLPSAV
YRKSGTVDSL DEIPPQNNNV PPRQGFSHRL SHVSMFRSGF SNSSVSIIRA
PMFSWIHRSA EFNNIIPSSQ ITQIPLTKST NLGSGTSVVK GPGFTGGDIL
RRTSPGQIST LRVNITAPLS QRYRVRIRYA STTNLQFHTS IDGRPINQGN
FSATMSSGSN LQSGSFRTVG FTTPFNFSNG SSVFTLSAHV FNSGNEVYID
RIEFVPAEVT

3/3

FIGURE 3

MNPNNRSEHD TIKTTENNEV PTNHVQYPLA ETPNPTLEDL NYKEFLRMTA
DNNTEALDSS TTKDVIQKGI SVVGDLLGVV GFPPGGALVS FYTNFLNTIW
PSEDPWKAFM EQVEALMDQK IADYAKNKAL AELQGLQNNV EDYVSALSSW
QKNPVSSRNP HSQGRIRELF SQAESHFRNS MPSFAISGYE VLFLTTYAQA
ANTHLFLLKD AQIYGEEWGY EKEDIAEFYK RQLKLTQEYT DHCVKWYNVG
LDKLRGSSYE SWVNFNRYRR EMTLTVLDLI ALFPLYDVRL YPKEVKTELT
RDVLTDPIVG VNNLRGYGTT FSNIENTYIRK PHLFDYLHRI QFHTRFQPGY
YGNSDFNYWS GNYVSTRPSI GSNDIITSPF YGNKSSEPVQ NLEFNGEKVY
RAVANTNLAV WPSAVYSGVT KVEFSQYNDO TDEASTQTYD SKRNVGAVSW
DSIDQLPPET TDEPLEKGYS HQLNYVMCFL MQGSRGTPIV LTWTHKSVDF
FNMIDSKKIT QLPLVKAYKL QSGASVVAGP RFTGGDIIQC TENGSAATYI
VTPDVSYSQK YRARIHYASR SQITFTLSLD GAPFNQYYFD KTINKGDTLT
YNSFNLASFS TPFELSGNNL QIGVTGLSAG DKVYIDKIEF IPVN

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
22 May 2003 (22.05.2003)

PCT

(10) International Publication Number
WO 2003/042656 A3

(51) International Patent Classification⁷: **C12P 21/00**,
G01N 33/53

(21) International Application Number:
PCT/US2002/034879

(22) International Filing Date: 30 October 2002 (30.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/350,650 13 November 2001 (13.11.2001) US

(71) Applicant (for all designated States except US): **GENEN-
COR INTERNATIONAL, INC.** [US/US]; 925 Page Mill
Road, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **HARDING, Fiona**,
A [US/US]; 772 Lewis Street, Santa Clara, CA 95050 (US).
POWER, Scott, D. [UG/UG]; 524 Poplar Avenue, San
Bruno, CA 94066 (US).

(74) Agent: **MACKNIGHT, Kamrin, T.**; GENENCOR IN-
TERNATIONAL, INC., 925 Page Mill Road, Palo Alto,
CA 94304 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK,
TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(88) Date of publication of the international search report:
4 March 2004

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.



WO 2003/042656 A3

(54) Title: IDENTIFYING AND REDUCING THE ALLERGENICITY OF FOOD PROTEINS

(57) Abstract: The present invention provides methods and composition suitable for the identification and modification of allergenic proteins. More particularly, the present invention provides methods and compositions for the identification and modification of epitopes of allergenic food proteins, as well as epitopes of potentially allergenic food proteins.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/34879

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12P 21/00; G01N 33/53

US CL : 435/7.24, 69.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/7.24, 69.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99/53038 A2 (GENENCOR INTERNATIONAL, INC.) 21 October 1999. See page 2, line 1-page 4, line 29; page 5, line 26-page 6, line 19 (especially page 6, lines 6-7).	1-12

Y		1-12
Y	WO 01/40281 A2 (THROMB-X) 07 June 2001. See page 3, line 19-page 5, line 22; page 7, line 20-page 9, line 3.	1-12
E, Y	US 2003/0041354 A1 (KJAERULFF et al) 27 February 2003. See page 1, paragraphs [0003]-[0009]; page 9, paragraph [0115]-page 11, paragraph [0135]; page 12 paragraph [0173]-page 13 paragraph [0175]; page 13, paragraph [0184].	1-12

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search

08 August 2003 (08.08.2003)

Date of mailing of the international search report

22 AUG 2003

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Facsimile No. (703)305-3230

Authorized officer

Valerie Bell-Hamilton
David A. Saunders, PhD

Telephone No. 703-308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

PCT/US02/34879

Continuation of B. FIELDS SEARCHED Item 3:
EAST:

(t adj cell adj epitop\$2) and (allerg\$10 near2 food)

((((reduc\$4 or modif\$7 or lower\$3 or abbrogat\$4 or less\$4) adj (allerg\$10 or hypersensiti\$10 or atopy or atopic or immunogenic\$5 or antigenic\$5)) and ((diet\$3 or ingest\$5 or cry\$3 or ber) near3 protein) and (t adj cell adj epitop\$2)

Form PCT/ISA/210 (second sheet) (July 1998)